CS310 : Automata Theory 2019

Lecture 10: Pumping lemma

Instructor: Ashutosh Gupta

IITB, India

Compile date: 2019-01-24

Non-regular languages

Are there languages that are not regular?

Yes!

Example: non-regular languages

Example 10.1

Consider the following language

$$L_{eq} = \{0^n 1^n | n \geq 0\}$$

Theorem 10.1

L_{eq} is not a regular language.

Proof.

Let us suppose L_{eq} is regular. So, there is a DFA A such that $L(A) = L_{eq}$.

Since A has finite states and due to the pigeonhole principle,
there must be different i and j such that 0^i and 0^j bring A to the same state.

...

Jargon alert: pigeonhole principle

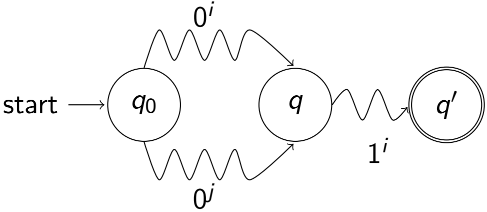
Pigeonhole principle states that if n items are put into m containers, with $n > m$, then at least one container must contain more than one item.

Example: non-regular languages II

Proof(contd.)

Afterwords, let us feed 1^i to the automaton.

Since the automaton is deterministic, 0^i1^i and 0^j1^i will land on same state.



Since $0^i1^i \in L_{eq}$, q' is accepting.

Therefore, $0^j1^i \in L(A)$. Contradiction. □

Proving some language is not regular

The argument in the previous theorem appears to be the language specific.

However, the argument can be generalized, which is called

pumping lemma.

Pumping lemma

Theorem 10.2

Let L be a regular language. Then there exists a constant n such that for every word $w \in L$ such that $|w| \geq n$, we can break w into three words $w = xyz$, such that

1. $y \neq \epsilon$,
2. $|xy| \leq n$, *and*
3. *for all* $k \geq 0$, $xy^k z \in L$.

Let us write the above lemma as a logical statement .

$$\exists n.\ \forall w \in L.\ (|w| \geq n \Rightarrow \exists xyz.\ (xyz = w \wedge y \neq \epsilon \wedge |xy| \leq n \wedge (\forall k.\ xy^k z \in L)))$$

Before proving this lemma let us understand the lemma.

Understanding pumping lemma

We call it pumping lemma because we can "pump" y any number of times including 0.

Exercise 10.1

How many quantifier alternations are there in the lemma?

Exercise 10.2

Show pumping lemma holds on the finite languages.

Exercise 10.3

For a given n, how many xyz breaks of w are possible?

Exercise 10.4

Why $y = \epsilon$ is disallowed?

Proving pumping lemma

Proof (of Theorem 10.2).

Let us suppose L is regular. Let A be a DFA such that $L = L(A)$.

Let us suppose A has n states.

Now consider a word $w = a_1 \dots a_m \in L$ such that $m \geq n$.

Let $p_0, \dots, p_m$ be the run of A on w. Therefore, p_m is an accepting state.

By the pigeonhole principle, at least two states in $\underbrace{p_0, \dots, p_n}_{n+1}$ should be equal.

Therefore, there are i and j such that $0 \leq i < j \leq n$ and $p_i = p_j$. ...

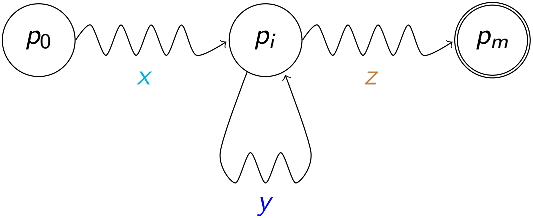
Proving pumping lemma II

Proof (contd.).

Let us propose the break of $w = xyz$ as follows.

- $x = a_1 a_2 ... a_i$
- $y = a_{i+1} a_{i+2} ... a_j$
- $z = a_{j+1} a_{j+2} ... a_m$

By construction, $y \neq \epsilon$ and $|xy| \leq n$.(why?)

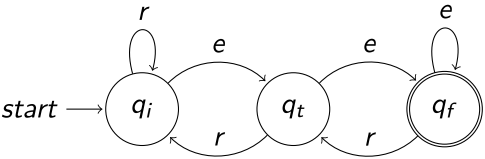


Now it is clear for any $k \geq 0$, xy^kz takes A to p_m, which is accepting. Therefore, $xy^kz \in L$. □

Example: pumping words

Example 10.2

Consider DFA



The size of DFA is $n = 3$*. Let us choose a word longer than n and accepting.*

$$w = ereere$$

The run on the word is $\underbrace{q_i q_t q_i q_t}_{n+1} q_f q_t q_f$.

We have q_t *repeated in the first four states. Therefore,*

- $x = e$
- $y = re$
- $z = ere$

Therefore, $e(re)^k ere \in L$.

Use of pumping lemma

Pumping lemma is very useful in proving that a language is not regular.

We apply contrapositive of the pumping lemma.

Recall the condition of pumping lemma.

$$\exists n.\ \forall w \in L.\ (|w| \geq n \Rightarrow \exists xyz.\ (xyz = w \wedge y \neq \epsilon \wedge |xy| \leq n \wedge (\forall k.\ xy^k z \in L)))$$

The negation of the above condition is

$$\forall n.\ \exists w \in L.\ (|w| \geq n \wedge \forall xyz.\ (xyz = w \wedge y \neq \epsilon \wedge |xy| \leq n \Rightarrow (\exists k.\ xy^k z \in L)))$$

Contrapositive of pumping lemma

We can turn the formula into words.

Theorem 10.3

Let L be a language. L is not regular if, for each n, there is a word $w \in L$ such that $|w| \geq n$ and for each breakup of w into three words $w = xyz$ such that

1. *$y \neq \epsilon$ and*
2. *$|xy| \leq n$,*

then there is a $k \geq 0$ such that $xy^kz \notin L$.

We will see how to apply the above lemma.

To be continued...

End of Lecture 10